

Constructing Learning Maps for Lecture Videos by Exploring Wikipedia Knowledge

Feng Wang¹, Xiaoyan Li¹, Wenqiang Lei¹, Chen Huang², Min Yin¹(✉),
and Ting-Chuen Pong³

¹ Shanghai Key Laboratory of Multidimensional Information Processing,
East China Normal University, 500 Dongchuan Road, Shanghai, China
myin@cs.ecnu.edu.cn

² School of Computer Science and Engineering, University of Electronic Science
and Technology of China, Chengdu, China

³ Department of Computer Science and Engineering, Hong Kong University
of Science and Technology, Hong Kong, China

Abstract. Videos are commonly used as course materials for e-learning. In most existing systems, the lecture videos are usually presented in a linear manner. Structuring the video corpus has proven an effective way for the learners to conveniently browse the video corpus and design their learning strategies. However, the content analysis of lecture videos is difficult due to the low recognition rate of speech and handwriting texts and the noisy information. In this paper, we explore the use of external domain knowledge from Wikipedia to construct learning maps for online learners. First, with the external knowledge, we filter the noisy texts extracted from videos to form a more precise and elegant representation of the video content. This facilitates us to construct a more accurate video map to represent the domain knowledge of the course. Second, by combining the video information and the external academic articles for the domain concepts, we construct a directed map to show the relationships between different concepts. This can facilitate online learners to design their learning strategies and search for the target concepts and related videos. Our experiments demonstrate that external domain knowledge can help organize the lecture video corpus and construct more comprehensive knowledge representations, which improves the learning experience of online learners.

1 Introduction

With the great advances in Web and multimedia technologies, e-learning has become one of the most important ways of learning new knowledge and skills. In the past few years, MOOC (Massive Open Online Course) has attracted numerous attentions from IT education and multimedia communities. MOOC provides services to learners by sharing educational resources through Internet. It is open accessible, scalable, and free to all social masses. This enables people to learn anywhere at anytime. However, due to the lack of instructions during learning, online learners have to face a mass of course materials and decide the

suitable learning strategies by themselves. Without any instructions, they may easily get lost in the mass of materials. Knowledge representation has proven an effective way for e-learning [1]. To improve the learning experience, in this paper, we address the construction of learning maps by structuring the course materials to facilitate more efficient browsing for learners.

Among all online course materials, video has been the most commonly used in the current e-learning platforms since it captures the vivid presentation of the domain knowledge from the instructors. In most existing systems, lecture videos are usually presented in a linear manner according to the lecture presentation and recording time. This can be used by learners to linearly browse the lecturer's presentations from the beginning to the end of the course. However, the inner structure of the domain knowledge of a course is usually not linear. For those people who just want to learn a specific domain concept, they prefer to watch only the videos with the related and the prerequisite concepts instead of the whole video corpus. Even for those learners who learn the whole course, a well-organized video corpus can help them to have an overview of the domain knowledge and plan their learning. Thus, it is necessary to re-organize the lecture video corpus to better represent the knowledge of the course so as to facilitate efficient and effective learning.

Topic threading of video corpus has been actively researched in multimedia field [7, 8]. In [8], news video archives are documented by story clustering and threading. To discover the relationship between videos, visual, speech, and text information are frequently employed. However, for lecture videos, visual information is usually less informative where only the lecturer and the white-board/slides are present with limited noticeable visual changes between different videos. Instead, the domain concepts that the videos present are the most informative which contain the semantic information of the presentation. In our previous work [3], we propose a framework to construct the knowledge representation by using speech transcript and handwriting texts which contain the domain concepts. However, it is nontrivial to extract the concept words from videos. The speech usually contains many noisy words besides the concepts. The handwriting texts are clearer; however, the recognition rate of handwriting is very low. This reduces the effectiveness of the semantic understanding of the presentations.

Besides video presentations, there are abundant domain knowledge about the courses available from other resources such as Web. In this paper, we explore the external domain knowledge from Wikipedia to aid the analysis of lecture videos. The external knowledge can help the structuring of lecture videos from the following two aspects. First, by identifying the domain concepts, we can filter the noisy information from videos including speech transcripts and handwriting texts. This results in a more elegant and precise representation of the lecture content. We then construct a video map to better represent the relationships between different videos (Sect. 2). Second, by combining videos and the external knowledge, we further construct a concept based knowledge representation to represent the relationships between different concepts and help the learners design their learning strategies (Sect. 3).

The contribution of this paper lies in the following aspects: (i) We propose to employ external domain knowledge to boost the content analysis of lecture videos and the video based knowledge representation. (ii) We propose a revised TF-IDF weighting approach for the similarity measure between lecture videos. (iii) We propose to combine video presentations and academic articles to construct a concept based knowledge representation for online learners.

2 Construction of Video Map by Exploring Domain Knowledge from Wikipedia

The input to our system is a lecture video sequence $V = \{v_1, v_2, \dots, v_m\}$ for a given course ordered according to the lecture presentation and recording time. Each video captures the lecturer's presentation for one or few domain concepts of the course. In most existing e-learning platforms, the videos are linearly presented to the learners for browsing. In this section, we construct a video map so as to represent the relationships between different videos which can better facilitate the learning and searching of desired videos and concepts by learners. First, we extract the concept words from videos to represent the content of each video by exploring the external domain knowledge from Wikipedia. Second, we revise the TF-IDF (Term Frequency-Inverse Document Frequency) algorithm to measure the similarities between videos based on the extracted concept words. Finally, a video map is constructed as the knowledge representation of the video corpus.

2.1 Extraction of Concept Words

To discover the relationships between different videos, the first step is to understand the content of each video. Since lecture videos are captured to present the domain knowledge of a course, the domain concepts are the most useful information to represent the video content. In [3], speech transcripts and handwriting texts are used to represent the video content. However, both of them suffer from low recognition accuracy. Furthermore, besides the concept words, there are many noisy texts which will affect the similarity measure between videos. To precisely identify the concept words in videos, we employ the domain knowledge of the course from Web to filter the speech transcripts and handwriting texts.

First, for a given online course, we collect the domain concept words from Wikipedia [9] to form a concept word list $C^{(W)} = \{c_1, c_2, \dots, c_N\}$. Second, for each video in the corpus, the lecturer's speech and handwriting texts are recognized [3], and then filtered by removing those words that are not included in $C^{(W)}$. This results in a more precise and elegant representation of the knowledge content of the video by focusing only on the domain concepts.

2.2 Representing Video Content with Revised TF-IDF

With the filtered concept set, we can measure the video similarity in a more precise way. To this end, the content of each video v_i is represented by a $N - dim$

concept vector $w_i = \{w_{i1}, w_{i2}, \dots, w_{iN}\}$ where w_{ij} is the weight of the concept c_j for v_i .

For the weighting of each concept word, TF-IDF (Term Frequency - Inverse Document Frequency) is the most used. However, compared with text documents, the lecture video corpus has its own feature, i.e. the sequence information in the temporal domain. The importance of a concept word in two videos or at two temporal points may be different. For instance, if a concept word c_j appears in a video v_i for the first time (c_j never appears before v_i), it means the lecturer begins to introduce and present this concept. In this case, c_j is the focus of this video and the weight of c_j for video v_i should be increased. Although the concept c_j may be referred when presenting other related concepts in the following videos, the first presence is usually the most important and should be highlighted. Based on this observation, we append the temporal information to TF-IDF by assigning different weights to a given concept according to the time of its presence. More specifically, if a concept c_j has appeared many times before video v_i , its weight for v_i will be reduced, which is defined as

$$w_{ij} = \frac{\alpha}{\alpha + F_{ij}} \cdot \frac{tf_{ij}}{T_i} \cdot \log \frac{m}{df_j} \quad (1)$$

where tf_{ij} is the term frequency of concept c_j present in video v_i , $T_i = \sum_{p=1}^N tf_{ip}$ is the total number of concept words present in video v_i , df_j is the number of videos where c_j is present, $F_{ij} = \sum_{q<i} tf_{iq}$ is the total frequency that c_j is present in all videos before v_i , and α is a smooth factor which is empirically set to be 5 in our experiments. If v_i is the first video where c_j is present, $F_{ij} = 0$. Otherwise, if v_j has already appeared many times before v_i , its importance become minor with the growing of F_{ij} . Finally, by normalizing the weights in Eq. 1 as

$$w'_{ij} = \frac{w_{ij}}{\sum_{j=1}^m w_{ij}} \quad (2)$$

we represent the content of video v_i with a concept vector $a_i = (w'_{i1}, w'_{i2}, \dots, w'_{iN})$.

2.3 Construction of Video Map by Maximum Spanning Tree

Given a lecture video corpus for a course, we first construct a graph G with the videos as the vertexes. The edge between every two vertexes is weighted with the similarity between the two corresponding videos. Here we employ Cosine similarity to calculate the weight between two videos as

$$sim_{cos}(v_i, v_k) = \frac{a_i \cdot a_k}{\|a_i\| \cdot \|a_k\|} \quad (3)$$

where a_i, a_k are the feature vectors for the two videos respectively calculated in Sect. 2.2.

Although the graph G presents the similarities between different videos, as a complete graph, it is less informative and thus learners cannot easily have an overview of the knowledge structure of the video corpus. Next, we simplify

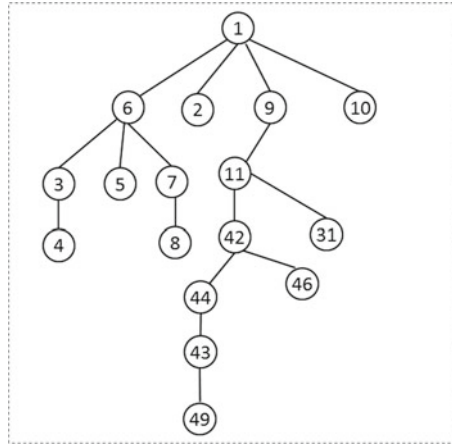


Fig. 1. Part of the video map for the course Chemistry.

the graph to be a tree where the edges with larger weights are preserved and the edges with small weights are removed. This is carried out by employing the Maximum Spanning Tree (MST) algorithm as in our previous work [3]. Figure 1 illustrates an example of the video map for the course *Chemistry*. By representing the video corpus with a tree, users can easily find the learning path by watching the related videos when they are going to explore a specific concept instead of linearly browsing all the videos.

3 Construction of Concept Map by Integrating Wikipedia Knowledge and Lecture Videos

The video based knowledge representation in Sect. 2 shows the relationship between different videos. Most time learners prefer to learn based on the relationships between different concepts and view the video corpus based on concepts. In this section, we construct a learning map between all domain concepts presented in the video corpus. Although videos contain explanations to the concepts, it is difficult to discover the relationship between different concepts based on videos. Each video may contain few different concepts and it is hard to segment a video into clips corresponding to different concepts in order to discover the relationship between them. In our approach, we employ the Wikipedia articles about the domain concepts to discover the relationships between different concepts and construct the learning maps between them. First, we construct an undirected graph to present the relationships between different concepts based on the domain articles. Second, with the lecture videos, we calculate the prerequisite relationship between every two semantically related concepts and generate a directed concept map. If c_a is a prerequisite concept of c_b , the learners need to learn c_a in order to learn c_b . This is represented in our concept by adding a

directed edge from c_a to c_b . By identifying the prerequisite relationship between different concepts, learners can easily design their learning strategies according to the concept map.

The construction of concept map has been researched in some previous works. In [5], concept map is generated to find Remedial-Instruction path so as to assist students learn better. The lecture needs to manually pre-set the weights and the threshold during the construction of the concept map. In [4], text information from academic articles is employed to build a domain concept map. The relation strength of two concepts is measured and PCA algorithm is used to construct the concept map. However, the prerequisite relationships between different concepts are not revealed. In [6], a concept map is constructed by comparing the weight between two concepts with a predefined threshold. The prerequisite relationships between different concepts are also not presented. In our approach, lecture videos and the Wikipedia articles are integrated so as to automatically discover the semantic relatedness and the prerequisite relationships between different concepts. More specifically, the academic articles contain the detailed semantic relationship between different concepts, while the video sequence implies the prerequisite relationships between them.

3.1 Constructing Undirected Concept Map with Wikipedia Articles

For each domain concept c_i , we download the corresponding article D_i from Wikipedia, which can be used as an explanatory document for the concept. Next, the semantic content of D_i is represented by an N -dimension vector $u_i = (u_{i1}, u_{i2}, \dots, u_{iN})$ where u_{ij} counts the presence of the concept c_j in D_i . Based on this representation, we calculate the semantic relatedness between every two concepts with Cosine similarity.

A graph \mathcal{G} is then constructed with each concept as a vertex. The edge e_{ab} between two concept c_a and c_b is weighted by the similarity between them. Next, we simplify \mathcal{G} by removing those edges with small weights. An edge e_{ab} is eliminated if its weight $s_{ab} < 3 \cdot \bar{s}$, where \bar{s} is the average weight of all edges in \mathcal{G} .

3.2 Constructing Directed Concept Map by Discovering the Prerequisite Relationships Between Concepts

The map constructed in Sect. 3.1 is undirected, which is less useful for learners to decide their learning strategies. In this section, we add directions to the edges to show the prerequisite relationship between two concepts. This is carried out by exploring the temporal information of the concepts presented in lecture videos.

Generally speaking, if c_j is a prerequisite concept of c_k , learners need to learn c_i before learning c_k . In the video corpus, c_j is usually presented before c_k . Here we calculate the prerequisite score for a concept c_j according to its presence time in the video corpus. This is achieved by using the weights calculated in Sect. 2.1 which measures both the temporal information and the importance of a concept in each video. For a concept c_j , we get a ranked video list L in the descending

order of the weights for c_j in different videos, i.e. for every p , $1 \leq p < m$, $w'_{L(p),j} > w'_{L(p+1),j}$. The prerequisite score for concept c_j is calculated as

$$Pr(j) = \sum_{p=1}^M \frac{L(p) \cdot w'_{L(p),j}}{M} \quad (4)$$

where $L(p)$ is the ordinal number in the video sequence V of the p -th video in L , and M is empirically set to 3 in our experiments. The prerequisite score for a concept c_j actually estimates the ordinal number of the video which presents c_j in the lecture video sequence.

Based on the prerequisite score, we add directions to the edges in the concept map. For each edge connecting two concepts c_j and c_k , we replace it with an directed edge $c_j \rightarrow c_k$ if $Pr(j) \leq (1 + \beta) \cdot Pr(k)$, where β is a smooth factor and set to be 0.05 in our experiment. For those remaining undirected edges, the two related concepts are parallel presented with a high probability. They can be regarded as a concept cluster with very close semantic relationship. By combining lecture videos and academic articles, we can discover both the semantic relatedness and the prerequisite relationship between different concepts which cannot be accomplished by using either single material.

Next, we remove the redundant dependencies in the concept map. An edge from concept c_j to c_k is considered as redundant if there exists another path from c_j to c_k and the length of the path is larger than 1. We traverse the concept map in a depth-first manner to find and remove all the redundant edges.

Finally, in the resulting concept map, we associate each concept c_j with the first three videos from the video list L . With the concept map, learners can view the concept based knowledge structure of the course. If they would like to learn a specific concept, they can find the prerequisite concepts and design the learning path according to the directed map. During learning, they can search a specific concept in the video corpus by clicking the videos associated to it instead of blindly clicking the linearly arranged videos. Figure 2 shows an example of the concept map where each node is a concept from the course *Biology*. For instance, if a learner would like to learn the concept 25, s/he can find its prerequisite concepts and learn them by watching the associated videos along the red paths as shown in Fig. 2.

4 Experiments

Our experiments are carried out on the lecture videos from Khan Academy [10]. Three video corpuses are used for the courses *Chemistry*, *Biology*, and *Physics* respectively. For external domain knowledge, we download the concept list for each course and the explanatory articles for each concept from Wikipedia [9]. After word lemmatization, a concept list is generated for each course. Table 1 shows the statistics of the video corpuses and the concepts for the three courses.

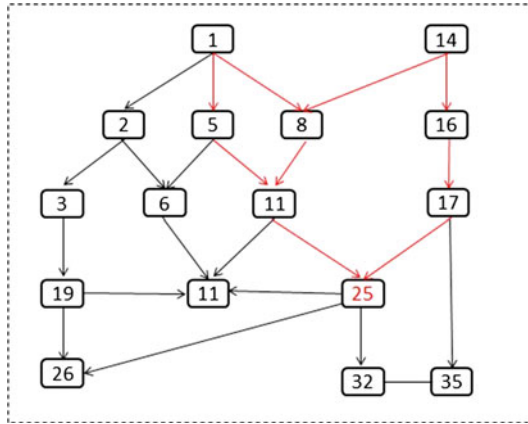


Fig. 2. Part of the concept map for the course Biology.

Table 1. The statistics of the video corpuses for three courses.

| Course | Number of videos | Number of concepts |
|-----------|------------------|--------------------|
| Chemistry | 103 | 240 |
| Biology | 72 | 72 |
| Physics | 201 | 223 |

4.1 Evaluation of Video Maps

In the first experiment, we evaluate and compare the video based knowledge representation between the approaches in this paper and in [3]. We ask the professors and students majored in the corresponding subject to score the resulting learning maps according to the correctness and comprehensiveness of the knowledge representation. Each judge gives a score ranging from 1 to 5 for the generated maps with 5 meaning the best. Table 2 shows the mean scores for different approaches and courses.

Overall, our video map gets higher scores than the approach in [3]. Most judges comment that the video maps by our approach look more accurate and comprehensive. This is because the external domain knowledge is employed in our approach which enables us to more precisely understand the lecture content and estimate the relationship between different videos. Besides video maps, we provide users with directed concept maps. By referring to both video maps and concept maps, learners can better view the knowledge structure of the course and more efficiently browse the target concepts and videos. Our experiment shows that the concept map is enjoyable for online learners and lecturers.

Table 2. Subjective Evaluation of the Different Approaches.

| Course | Video map in [3] | Video map with our approach | Video map + Concept map |
|-----------|------------------|-----------------------------|-------------------------|
| Chemistry | 3.64 | 4.25 | 4.59 |
| Biology | 3.87 | 4.33 | 4.63 |
| Physics | 3.41 | 4.19 | 4.42 |

4.2 Evaluation of Concept Maps

In the second experiment, we evaluate the directed concept maps generated in Sect. 3. The resulting concept maps are evaluated from two aspects: the accuracies of the prerequisite relationship (or the directed edges) and the concept-video association. This is carried out by asking the judges to manually label each directed edge in the concept map and the videos associated to each concept.

Table 3 shows the evaluation results for the prerequisite relationship in the concept map. For each concept, we can find one or few prerequisite concepts. The accuracies of the directed edges are quite encouraging. The judges agree that the concept maps can help a lot as instructions for online learners. Our experiments show that integrating lecture videos and external domain knowledge is a promising way of constructing effective knowledge representation.

Table 3. Accuracies of the prerequisite relationship in the concept map.

| Course | # directed edges | # corrected directed edges | Accuracy |
|-----------|------------------|----------------------------|----------|
| Chemistry | 544 | 488 | 89.7 % |
| Biology | 195 | 184 | 94.4 % |
| Physics | 514 | 472 | 91.8 % |

Table 4 presents the evaluation results of the concept-video association. In the concept map, we associated each node with three videos which present the corresponding concept so that learners can easily find and browse the lecturer's presentations. As shows in Table 4, for most concepts, the first video we provide to learners is the primary presentation by the lecturer. For the remaining concepts, the corresponding presentations can also be found in the second or the third videos. Some concepts are presented in more than one videos and thus we associate each concept with three videos. The concept-video association is mainly based on the weighting of the concept in videos with the revised TF-IDF approach in Sect. 2.1. This experiment validates the effectiveness of our proposed approach.

Table 4. Accuracies of the concept-video association (%).

| Course | The first video | The second video | The third video |
|-----------|-----------------|------------------|-----------------|
| Chemistry | 82.9 | 16.7 | 6.3 |
| Biology | 90.3 | 11.1 | 2.8 |
| Physics | 87.4 | 13.5 | 8.1 |

5 Conclusions

We have presented our approach for constructing knowledge representations of online lecture videos. In our approach, two learning maps are provided to online learners as instructions for them to more efficiently browse the course materials and design their learning strategies. Compared with previous works, we explore the external domain knowledge from Wikipedia and integrate it with the recorded presentations from lecturers. This enables us to more precisely understand the semantic content of lecture videos in constructing comprehensive knowledge representations for the courses. Our experiments demonstrate that the resulting learning maps are accurate, comprehensive, and enjoyable by the online learners. For future work, we will explore how to better organize lecture videos and materials from different sources and present them to learners so as to make online learning more efficient, effective, and enjoyable.

Acknowledgments. The work described in this paper was supported by the Science and Technology Commission of Shanghai Municipality under research grant no. 14DZ2260800, the National Natural Science Foundation of China (No. 1103127, No. 61375016), SRF for ROCS, SEM, and the Fundamental Research Funds for the Central Universities.

References

1. Lee, J., Segev, A.: Knowledge maps for e-learning. *Comput. Educ.* **59**(2), 353–364 (2012)
2. Shaw, B.: A study of learning performance of e-learning materials design with knowledge maps. *Comput. Educ.* **54**(1), 253–264 (2010)
3. Fan, P.M., Pong, T.C.: Constructing knowledge representation from lecture videos through multimodal analysis. *Int. J. Inf. Educ. Technol.* **3**(3), 304–309 (2013)
4. Chen, N.S., Kinshuk, Wei, C.W., Chen, H.J.: Mining e-learning domain concept map from academic articles. *Comput. Educ.* **50**(3), 1009–1021 (2008)
5. Lee, C.H., Lee, G.G., Leu, Y.: Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning. *Expert Syst. Appl.* **36**(2), 1675–1684 (2009)
6. Tseng, Y.H., Chang, C.Y., Chang, S.N., Rundgren, C.J.: Mining concept maps from news stories for measuring civic scientific literacy in media. *Comput. Educ.* **55**(1), 165–177 (2010)

7. Pang, L., Zhang, W., Ngo, C.W.: Video hyperlinking: libraries and tools for threading and visualizing large video collection. In: ACM Multimedia Conference (2012)
8. Wu, X., Ngo, C.W., Li, Q.: Threading and autodocumenting in news videos. IEEE Signal Process. Mag. **23**(2), 59–68 (2006)
9. Web Concept Glossary from Wikipedia. http://en.wikipedia.org/wiki/Glossary_of_chemistry_terms, http://en.wikipedia.org/wiki/Glossary_of_biology, http://en.wikipedia.org/wiki/Glossary_of_physics
10. Khan Academy. <https://www.khanacademy.org/>